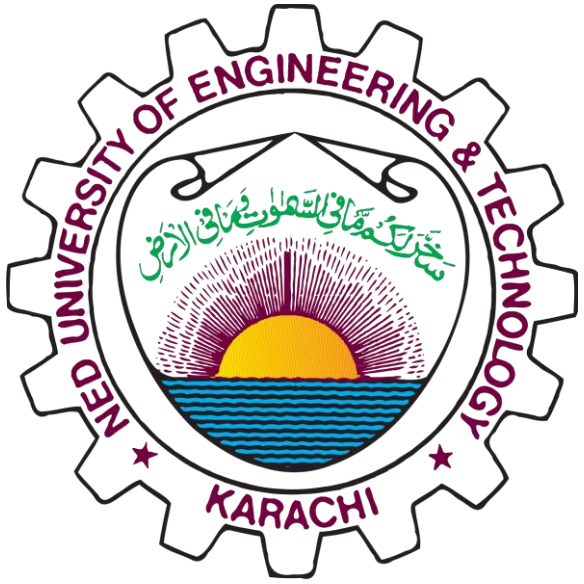


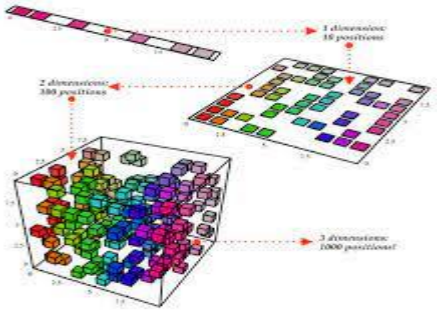
CT-562
MACHINE LEARNING

NED University of Engineering & Technology

DIMENSIONALITY REDUCTION AND CLASSIFICATION

DIMENSIONALITY REDUCTION

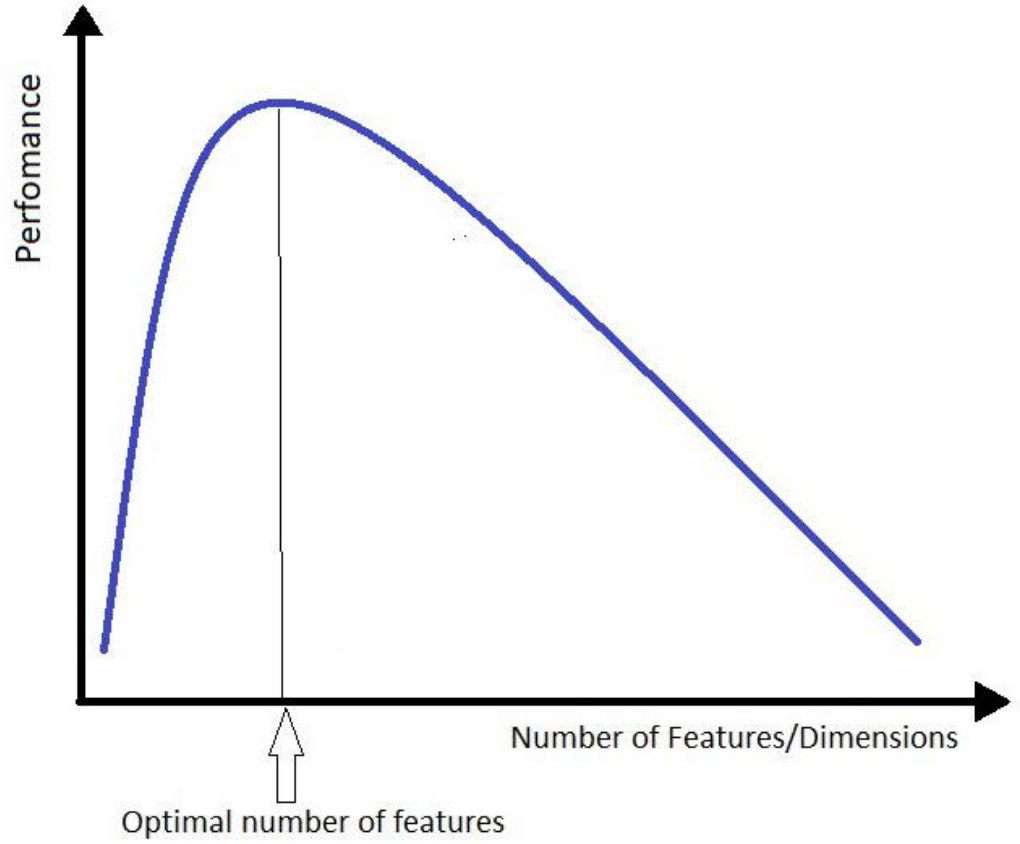
- The number of input features, variables, or columns present in a given dataset is known as dimensionality, and the process to reduce these features is called dimensionality reduction.



WHY DIMENSIONALITY REDUCTION?

- A dataset contains a huge number of input features in various cases, which makes the predictive modeling task more complicated. Because it is very difficult to visualize or make predictions for the training dataset with a high number of features, for such cases, dimensionality reduction techniques are required to use.
- Dimensionality reduction technique can be defined as, ***"It is a way of converting the higher dimensions dataset into lesser dimensions dataset ensuring that it provides similar information."*** These techniques are widely used in machine learning for obtaining a better fit predictive model while solving the classification and regression problems.

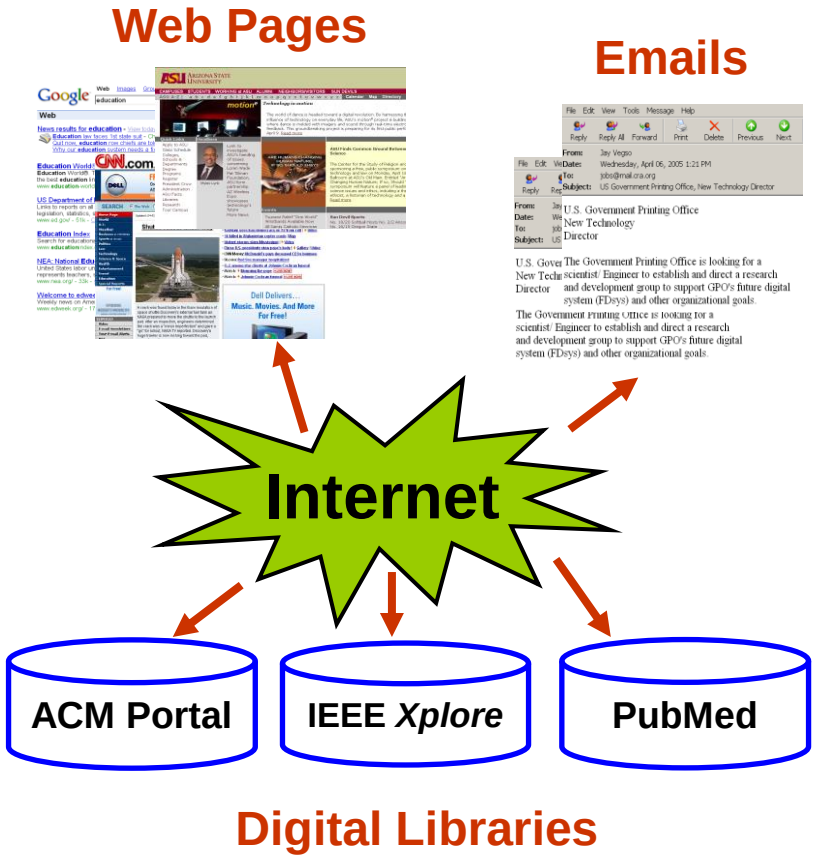
DIMENSIONALITY



WHY DIMENSIONALITY REDUCTION?

- Visualization: projection of high-dimensional data onto 2D or 3D.
- Data compression: efficient storage and retrieval.
- Noise removal: positive effect on query accuracy.

DOCUMENT CLASSIFICATION

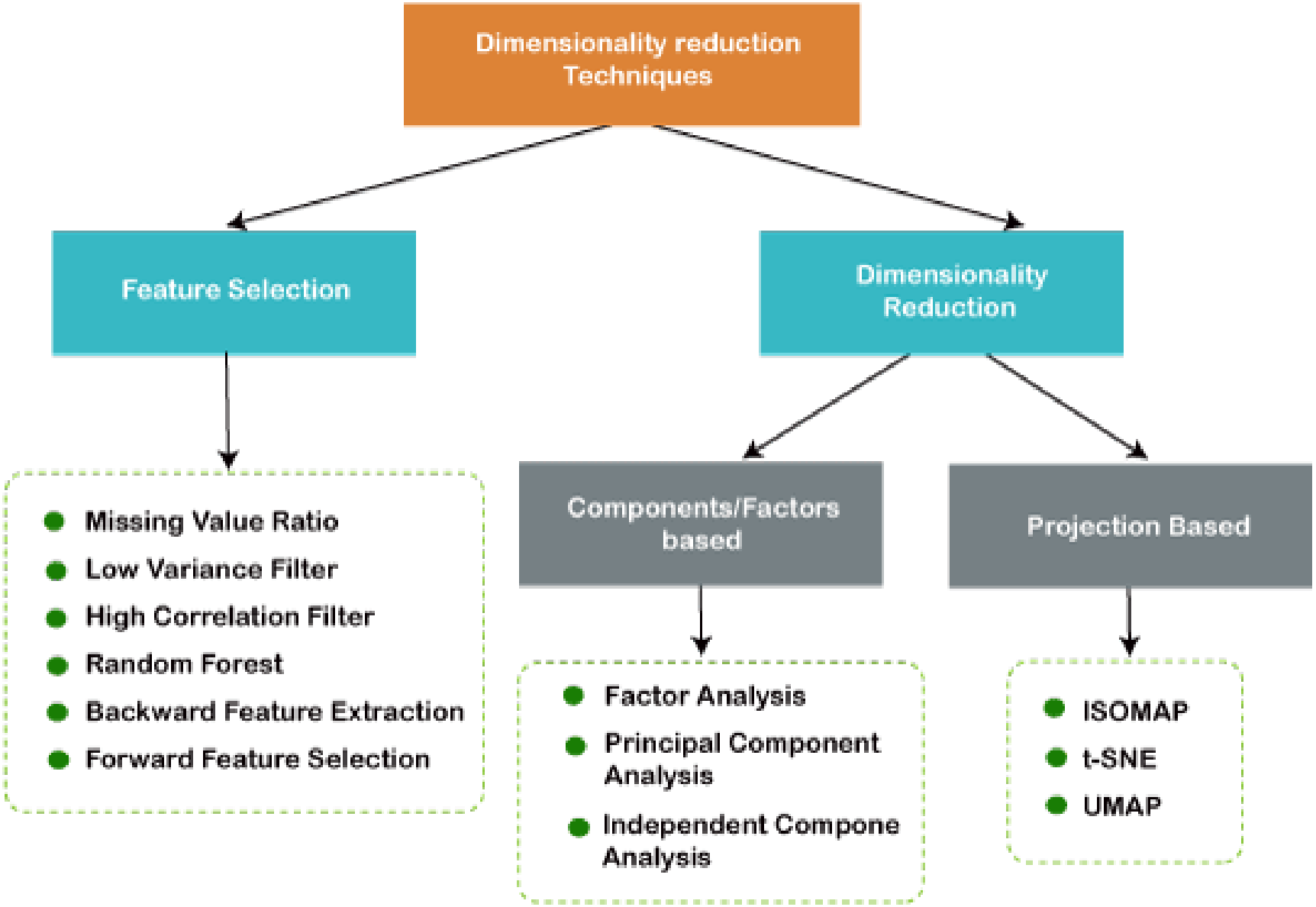


Documents	Terms T_1	T_2	.....	T_N	C
D_1	12	0	.....	6	Sports
D_2	3	10	.....	28	Travel
⋮	⋮			⋮	⋮
D_M	0	11	.....	16	Jobs

- **Task:** To classify unlabeled documents into categories
- **Challenge:** thousands of terms
- **Solution:** to apply dimensionality reduction

DIMENSIONALITY REDUCTION

- It is commonly used in the fields that deal with high-dimensional data, such as speech recognition, signal processing, bioinformatics, etc. It can also be used for data visualization, noise reduction, cluster analysis, etc.


Dimensionality reduction Techniques
Feature Selection
Dimensionality Reduction
Missing Value Ratio
Low Variance Filter
High Correlation Filter
Random Forest
Backward Feature Extraction
Forward Feature Selection
Components/Factors based
Projection Based
Factor Analysis
Principal Component Analysis
Independent Compone Analysis
ISOMAP
t-SNE
UMAP

THE CURSE OF DIMENSIONALITY

- Handling the high-dimensional data is very difficult in practice, commonly known as the curse of dimensionality. If the dimensionality of the input dataset increases, any machine learning algorithm and model becomes more complex. As the number of features increases, the number of samples also gets increased proportionally, and the chance of overfitting also increases. If the machine learning model is trained on high-dimensional data, it becomes overfitted and results in poor performance.

BENEFITS OF APPLYING DIMENSIONALITY REDUCTION

Some benefits of applying dimensionality reduction technique to the given dataset are given below:

- By reducing the dimensions of the features, the space required to store the dataset also gets reduced.
- Less Computation training time is required for reduced dimensions of features.
- Reduced dimensions of features of the dataset help in visualizing the data quickly.
- It removes the redundant features (if present) by taking care of multicollinearity.

DISADVANTAGES OF DIMENSIONALITY REDUCTION

There are also some disadvantages of applying the dimensionality reduction, which are given below:

- Some data may be lost due to dimensionality reduction.
- In the PCA dimensionality reduction technique, sometimes the principal components required to consider are unknown.

APPROACHES OF DIMENSION REDUCTION

- Feature Selection
- Feature Extraction

FEATURE SELECTION

Feature selection is the process of selecting the subset of the relevant features and leaving out the irrelevant features present in a dataset to build a model of high accuracy. In other words, it is a way of selecting the optimal features from the input dataset.

FEATURE SELECTION – METHODS

- Filters Methods
- Wrappers Methods

FILTERS METHODS

In this method, the dataset is filtered, and a subset that contains only the relevant features is taken. Some common techniques of filters method are:

- Correlation
- Chi-Square Test
- ANOVA

WRAPPERS METHODS

The wrapper method has the same goal as the filter method, but it takes a machine learning model for its evaluation. In this method, some features are fed to the ML model, and evaluate the performance. The performance decides whether to add those features or remove to increase the accuracy of the model. This method is more accurate than the filtering method but complex to work. Some common techniques of wrapper methods are:

- Forward Selection
- Backward Selection
- Bi-directional Elimination

2. WRAPPER METHODS (FEATURE SELECTION)

- **A. Forward Selection**
  - Forward Selection is an iterative method.
  - In this method, we start with one feature and we keep on adding features until no improvement in the model is observed.
  - The search is stopped after a pre-set criteria is met.
  - This is a greedy approach because it always targets the features in a forward fashion, which gives a boost to the performance.
  - If the number of features are large, it can be computationally expensive.
- **B. Backward Elimination**
  - This process is the opposite of the Forward Selection Method.
  - It starts initially with all the features and keeps on removing features until no improvement is observed.

WRAPPER METHOD: FORWARD FEATURE SELECTION

Steps to perform Forward Feature Selection

1. Train n model using each feature (n) individually and check the performance
2. Choose the variable which gives the best performance
3. Repeat the process and add one variable at a time
4. Variable producing the highest improvement is retained
5. Repeat the entire process until there is no significant improvement in the model's performance

WRAPPER METHOD: FORWARD FEATURE SELECTION

- Fitness level prediction

- So the first step in Forward Feature Selection is to train models using each feature individually and checking the performance.
- If you have three independent variables, we will train three models using each of these three features individually.

ID	Calories_burnt	Gender	Plays_Sport?	Fitness_Level
1	121	M	Yes	Fit
2	230	M	No	Fit
3	342	F	No	Unfit
4	70	M	Yes	Fit
5	278	F	Yes	Unfit
6	146	M	Yes	Fit
7	168	F	No	Unfit
8	231	F	Yes	Fit
9	150	M	No	Fit
10	190	F	No	Fit

FORWARD FEATURE SELECTION: EXAMPLE

- Let's say we trained the model using the **Calories_Burnt** feature and the target variable, **Fitness_Level** and we've got an accuracy of **87%**

ID	Calories_burnt	Gender	Plays_Sport?	Fitness_Level
1	121	M	Yes	Fit
2	230	M	No	Fit
3	342	F	No	Unfit
4	70	M	Yes	Fit
5	278	F	Yes	Unfit
6	146	M	Yes	Fit
7	168	F	No	Unfit
8	231	F	Yes	Fit
9	150	M	No	Fit
10	190	F	No	Fit

Accuracy = 87%

FORWARD FEATURE SELECTION: EXAMPLE CONT.

Next, we'll train the model using the **Gender** feature, and we get an accuracy of **80%**

Variable used	Accuracy
Calories_bumt	87.00%
Gender	80.00%
Plays_Sport?	85.00%

ID	Calories_bumt	Gender	Plays_Sport?	Fitness_Level
1	121	M	Yes	Fit
2	230	M	No	Fit
3	342	F	No	Unfit
4	70	M	Yes	Fit
5	278	F	Yes	Unfit
6	146	M	Yes	Fit
7	168	F	No	Unfit
8	231	F	Yes	Fit
9	150	M	No	Fit
10	190	F	No	Fit

Accuracy = 80%

EXAMPLE CONT.

- Next, we will repeat this process and add one variable at a time. So of course we'll keep the **Calories_Burnt** variable and keep adding one variable. So let's take **Gender** here and using this we get an accuracy of **88%-**

ID	Calories_burnt	Gender	Plays_Sport?	Fitness_Level
1	121	M	Yes	Fit
2	230	M	No	Fit
3	342	F	No	Unfit
4	70	M	Yes	Fit
5	278	F	Yes	Unfit
6	146	M	Yes	Fit
7	168	F	No	Unfit
8	231	F	Yes	Fit
9	150	M	No	Fit
10	190	F	No	Fit

Accuracy = 88%

EXAMPLE CONT

Plays_Sport along with **Calories_Burnt**, we get an accuracy of **91%.** A variable that produces the highest improvement will be retained.

ID	Calories_burnt	Gender	Plays_Sport?	Fitness_Level
1	121	M	Yes	Fit
2	230	M	No	Fit
3	342	F	No	Unfit
4	70	M	Yes	Fit
5	278	F	Yes	Unfit
6	146	M	Yes	Fit
7	168	F	No	Unfit
8	231	F	Yes	Fit
9	150	M	No	Fit
10	190	F	No	Fit

Accuracy = 91%

BACKWARD FEATURE ELIMINATION: EXAMPLE

- Fitness prediction level
- The first step is to train the model, using all the variables.
- You'll of course not take the ID variable train the model as ID contains a unique value for each observation
- So we'll first train the model using the other three independent variables. And of course, the target variable, which is the **Fitness_Level**.
- we get an **accuracy of 92% using all three independent variables.**

ID	Calories_burnt	Gender	Plays_Sport?	Fitness_Level
1	121	M	Yes	Fit
2	230	M	No	Fit
3	342	F	No	Unfit
4	70	M	Yes	Fit
5	278	F	Yes	Unfit
6	146	M	Yes	Fit
7	168	F	No	Unfit
8	231	F	Yes	Fit
9	150	M	No	Fit
10	190	F	No	Fit

Variable_dropped	Accuracy
Calories_burnt	90%

ID	Calories_burnt	Gender	Plays_Sport?	Fitness_Level
1	121	M	Yes	Fit
2	230	M	No	Fit
3	342	F	No	Unfit
4	70	M	Yes	Fit
5	278	F	Yes	Unfit
6	146	M	Yes	Fit
7	168	F	No	Unfit
8	231	F	Yes	Fit
9	150	M	No	Fit
10	190	F	No	Fit

Variable_dropped	Accuracy
Gender	91.60%

ID	Calories_burnt	Gender	Plays_Sport?	Fitness_Level
1	121	M	Yes	Fit
2	230	M	No	Fit
3	342	F	No	Unfit
4	70	M	Yes	Fit
5	278	F	Yes	Unfit
6	146	M	Yes	Fit
7	168	F	No	Unfit
8	231	F	Yes	Fit
9	150	M	No	Fit
10	190	F	No	Fit

Variable_dropped	Accuracy
Plays_Sport?	88%

ID	Calories_burnt	Gender	Plays_Sport?	Fitness_Level
1	121	M	Yes	Fit
2	230	M	No	Fit
3	342	F	No	Unfit
4	70	M	Yes	Fit
5	278	F	Yes	Unfit
6	146	M	Yes	Fit
7	168	F	No	Unfit
8	231	F	Yes	Fit
9	150	M	No	Fit
10	190	F	No	Fit

BACKWARD FEATURE ELIMINATION: EXAMPLE

- If you see gender has produced the smallest change in the performance in the model first, it was 92% when we took all the variables and when we dropped gender, it was 91.6%. So we can infer that gender does not have a high impact on the Fitness_Level variable. And hence it can be dropped.
- Finally, we will repeat all these steps until no more variables can be dropped.
- It's a very simple, but very effective technique.

Accuracy using all the variables = 92%

Variable_dropped	Accuracy
Calories_burnt	90%
Gender	91.60%
Plays_Sport?	88%

FEATURE EXTRACTION

Feature extraction is the process of transforming the space containing many dimensions into space with fewer dimensions. This approach is useful when we want to keep the whole information but use fewer resources while processing the information.

Some common feature extraction techniques are:

1. Principal Component Analysis
2. Linear Discriminant Analysis
3. Kernel PCA
4. Quadratic Discriminant Analysis

THANK YOU